EXHIBIT 2

The Evolving Quality of Groupon Deals (Draft)

Georgios Zervas
<zg@cs.yale.edu>
Computer Science Dept.
Yale University

May 14, 2012

Abstract

In this brief note, we present our preliminary findings on the evolving quality of Groupon deals as seen through the lens of the Yelp ratings of the merchants who offer them. Using simple regression analyses on a dataset spanning over two years of Groupon deals, we find that, on average, the Yelp ratings of merchants offering them has been on a downward slope.

1 Introduction

Groupon is a multi-billion dollar enterprise that has been growing rapidly since its inception in 2008. In this work, we ask a question that has implications for the sustainability of the daily deals business model: *as Groupon gets bigger, is it also getting better?* By *bigger*, we mean the growing volume and variety of deals that Groupon offers daily; by *better*, we mean the quality of these deals. The purpose of this brief note is to establish a few basic facts regarding the evolving quality of the deals Groupon offers.

Ideally, any measure of deal quality should capture a deal's consumer appeal, and hence its ability to generate sales. A good such measure is *conversion rate*: the fraction of consumers who purchase a deal over the number of consumers who are exposed to it. Unfortunately, even though Groupon publishes approximate sales data, we lack data on deal *impressions* (i.e., the number of times each deal is shown to consumers.) Instead, we resort to a proxy measure

which captures one aspect of consumer appeal: the Yelp rating of the merchant offering a deal at the time the deal starts. We justify this choice based on the following observations. First, since Yelp ratings are provided by consumers rather than by experts, they can be seen as a reflection of the population's opinion of a specific merchant. Second, Groupon and Yelp, to some extent, target the same audience of tech-savvy consumers; Groupon prominently displays Yelp ratings for many of the deals it offers, and, consumers can use these ratings to make purchase decisions. Third, previous work demonstrates that consumers use Yelp ratings as a signal for quality, at least as far as restaurants are concerned. In fact, changes in Yelp ratings have been shown to be causally associated with changes in both revenues and table reservations (Anderson & Magruder, 2011; Luca, 2011).

It is important to draw a distinction between a merchant's Yelp rating at the time a deal is offered, and subsequent to the time it is offered. Our focus here is on the measuring the former. Our previous work has focused on the latter; one of our findings was that the Yelp ratings of Groupon-bearing consumers were on average 10% lower than those of their peers (Byers *et al.* , 2012a,b).

We proceed by first describing our data collection process, and then by exploring some of the key features of our dataset. We then fit some simple regression models to the data, and conclude with a brief discussion of our results.

2 Data Description

Our data collection process proceeded in three steps. We began by using the Groupon API[1] to collect information on the complete set of deals that Groupon offered in four metropolitan areas between January 1st, 2010 and March 31st, 2012. Specifically, we collected information on 3,719 deals in Boston, 5,362 deals in New York, 3,764 deals in San Francisco, and 3,506 deals in Seattle. From these, we excluded Groupon Now!, Getaways, and Goods deals, as well as any deals not associated with a geographic location, such as those that could only be redeemed online. This left us with 2,892 deals in Boston, 4,230 in New York, 2,963 in San Francisco, and 2,695 in Seattle.

We then proceeded to map the merchants offering these deals to their corresponding Yelp pages. The mapping from deals to Yelp merchant pages is one-to-many: occcasionally, a Groupon deal can be redeemed at multiple locations of the same merchant, which in turn correspond to multiple Yelp pages. We associate Groupon deals to Yelp pages using two

[1] `http://www.groupon.com/pages/api`

methods. First, for about one third of the deals in our dataset, the Groupon API supplied us with a one or more links to the corresponding merchant's Yelp page(s). Second, for the remaining two thirds of the deals that the Groupon API supplied no Yelp link for, we used the Yelp API[2] to search for the merchants using their names and exact street addresses. The Yelp API responds to such queries with an ordered list of merchants names and locations, ranked by relevance (in a manner similar to which a user query on the Yelp website would work.) In most cases, the top search result provided the correct association. To guard against misassociations we employed both manual inspection of the results, as well as a set of heuristics (e.g., ensuring that the ZIP code of the merchant reported by Groupon matched that reported by Yelp.) Despite taking care to avoid errors, the possibility of wrong associations remains. At the conclusion of the process we had associated 2,432 Boston deals, 3,826 New York deals, 3,029, San Francisco deals, and 2,446 Seattle deals with their Yelp pages.

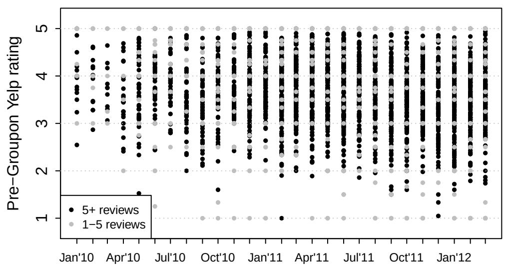
The final step in our data collection process involved gathering the complete review histories of the Groupon merchants that we succesfully associated with a Yelp page. To achieve this we crawled the Yelp website. In total, we collected 103,605 reviews for 1,569 Boston merchants, 192,522 for 2,533 New York merchants, 309,889 reviews for 2,086 San Francisco merchants, and 98,358 reviews for 1,496 Seattle merchants. The number of distinct merchants is smaller than the number of deals as merchants occasionally offer multiple Groupon deals. Some merchants had received zero reviews prior to their first Groupon deal.

3 Deal Quality over Time

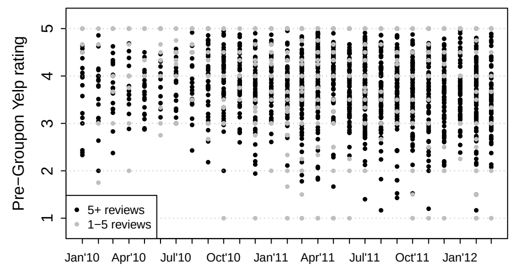
We investigate the relationship between deal quality, volume, and time, both diagrammatically, and by fitting simple regression models.

Figure 1 summarizes our dataset. Each data point in the four scatter plots corresponds to a single deal. Deals are grouped on the x-axis by the month they were offered. On the y-axis we plot the average Yelp rating of the merchant who offered the corresponding deal at the time the deal started. We further break down deals into two groups: deals by merchants with at least 5 reviews at the time the deal started are plotted in black; those with fewer than 5 reviews are plotted in grey. The choice of 5 reviews as a boundary is arbitrary but, nevertheless, draws a meaningful visual distinction in the amount of confidence we can have

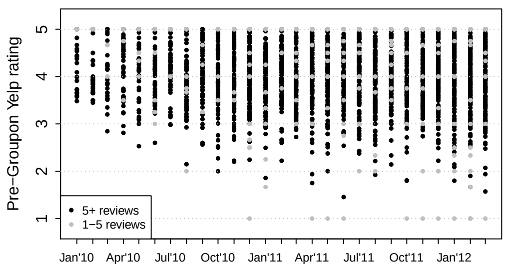
[2] `http://www.yelp.com/developers/documentation`



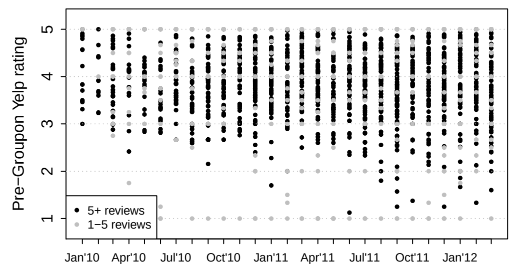
(a) New York



(b) Boston



(c) San Francisco



(d) Seattle

Figure 1: Yelp ratings of merchants at the time they offered Groupon deals.

in each deal's Yelp rating. Figure 1 reveals a trend: as the number of deals on offer has increased with time, so has the frequency of deals with Yelp ratings in the bottom half of the ratings scale (i.e., below 3 stars.) We proceed to examine this trend in more detail.

3.1 Mean Yelp rating over time

Figure 2 presents a monthly aggregation of the data. The vertical bars running across the bottom of the four subfigures indicate the number of deals offered per month. The grey lines display the average Yelp rating of all deals that ran each month. Linear trends are shown with black, dashed lines. A visual inspection suggests that the slope of the trend line is negative in all four cases; indicating Groupon deal quality is, on average, decreasing over time.

To rigorously measure the magnitude, and statistical significance of the trend we fit a linear regression model to our data. We associate each deal i with two quantities: y_i, the Yelp rating of the merchant offering the deal, at time the deal was offered; and, τ_i, a fractional time offset which is defined as the number of days between a deal's start date and 1/1/2012 in units of 30 days. This scaling has no consequence outside making the interpretation of

the estimated coefficients more convenient. We then fit the following model:

$$y_i = \beta_0 + \beta_1 \tau_i + \epsilon_i, \tag{1}$$

where ϵ_i is a normally distributed error term. The coefficient estimates, along with their t-values, are displayed in Table 1 under the "All deals" heading. In all four cases, the correlation between time and expected Yelp rating is negative, and statistically significant. The values of the coefficients are to be interpreted as changes in Yelp ratings every 30 days. For example, looking at the coefficient for Boston, we see an decrease in the expected Yelp rating of the deals on offer of 0.012 stars every 30 days.

Table 1 also reports a series of robustness checks. One concern is that the initial estimates do not take into account the varying number of reviews based on which each merchant's average Yelp rating is calculated. As a first check, we fit the model of Equation 1 to the subset of merchants that had at least 5 reviews at the time they offered their corresponding deals. The coefficient estimates remain largely unchanged with the exception of Seattle where the relationship becomes statistically insignificant.

Another concern is with merchants who have run multiple deals. Previous work has shown that Yelp ratings, on average, decline subsequent to Groupon deals Byers *et al.* (2012a,b). Therefore, the observed trend could be solely due to merchants who have run multiple deals since, at the time of subsequent deals, their Yelp rating could be lower. Thus, as a second robustness check we fit our model to the subset of deals that correspond to the first deal each merchant in our dataset ran. The largest difference is seen for Boston where the magnitude of the coefficient decreases significantly. The rest of the coefficients remain largely unchanged.

Our last robustness check simultaneously addresses both of the above concerns by looking at the first deal of each merchant who had at least 5 reviews at the time the deal was offered. Again, there is little change in the magnitude of the coefficients, and they remain statistically significant at the 5% level.

3.2 Frequency of low-rated deals over time

When considering deal quality, modeling the mean rating of the deals is just one way to look at the data. A shortcoming of this approach is that it doesn't inform us on how the underlying distribution of ratings varies over time. As an alternative we can look at the frequency of deals whose Yelp rating is 3 or fewer stars. For convenience, let us call these *low-rated deals*. Figure 3 displays the month-over-month percentage of low-rated deals, as

Table 1: Linear regression estimates of time offset on deal Yelp rating.

	(1) Boston	(2) New York	(3) San Francisco	(4) Seattle
All deals				
τ	-0.0120^{***}	-0.0139^{***}	-0.00571^{**}	-0.0110^{***}
	(-4.23)	(-6.16)	(-2.95)	(-3.85)
Constant	4.006^{***}	3.951^{***}	4.135^{***}	4.061^{***}
	(76.88)	(94.68)	(117.05)	(80.45)
Observations	2179	3341	2881	2116
At least 5 reviews at time of offer				
τ	-0.00958^{***}	-0.0134^{***}	-0.00569^{**}	-0.00466
	(-3.51)	(-6.20)	(-3.03)	(-1.78)
Constant	3.892^{***}	3.865^{***}	4.098^{***}	3.920^{***}
	(77.14)	(96.06)	(120.62)	(85.30)
Observations	1429	2421	2504	1386
First deals				
τ	-0.00729^{*}	-0.0122^{***}	-0.00474^{*}	-0.00951^{**}
	(-2.12)	(-4.49)	(-2.04)	(-2.61)
Constant	3.999^{***}	3.956^{***}	4.148^{***}	4.053^{***}
	(67.01)	(83.16)	(104.68)	(68.63)
Observations	1422	2245	2000	1329
First deals, at least 5 reviews at time of offer				
τ	-0.00705^{*}	-0.0130^{***}	-0.00562^{*}	-0.00804^{*}
	(-2.10)	(-5.03)	(-2.50)	(-2.37)
Constant	3.893^{***}	3.881^{***}	4.115^{***}	3.952^{***}
	(67.65)	(85.76)	(109.25)	(74.58)
Observations	872	1535	1672	808

t statistics in parentheses

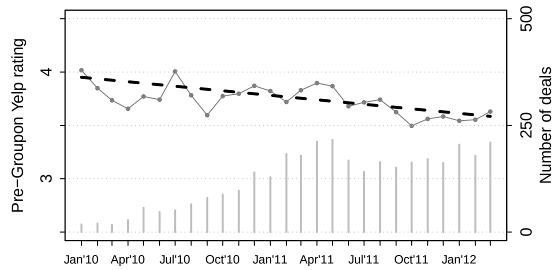
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Logistic regression estimates of time offset on likelihood of a low-rated deal. Dependent variable: Yelp rating of deal is at most 3 stars.

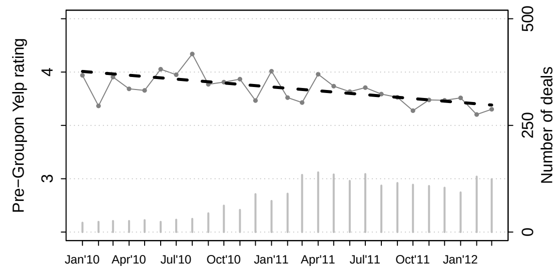
	(1) Boston	(2) New York	(3) San Francisco	(4) Seattle
All deals				
τ	0.0312***	0.0359***	0.0465***	0.0375***
	(3.45)	(5.10)	(4.47)	(4.17)
Constant	−2.090***	−2.028***	−3.219***	−2.297***
	(−12.10)	(−14.95)	(−15.70)	(−13.59)
Observations	2179	3341	2881	2116
At least 5 reviews at time of offer				
τ	0.0375**	0.0421***	0.0428***	0.0246*
	(3.09)	(4.85)	(3.60)	(2.08)
Constant	−2.444***	−2.250***	−3.322***	−2.362***
	(−10.39)	(−13.27)	(−14.32)	(−10.85)
Observations	1429	2421	2504	1386
First deals				
τ	0.0317**	0.0305***	0.0552***	0.0510***
	(2.89)	(3.68)	(4.55)	(4.46)
Constant	−2.151***	−1.955***	−3.342***	−2.514***
	(−10.81)	(−12.89)	(−14.48)	(−12.45)
Observations	1422	2245	2000	1329
First deals, at least 5 reviews at time of offer				
τ	0.0349*	0.0346**	0.0486***	0.0497**
	(2.33)	(3.25)	(3.42)	(3.05)
Constant	−2.443***	−2.166***	−3.412***	−2.803***
	(−8.98)	(−11.13)	(−12.95)	(−10.00)
Observations	872	1535	1672	808

t statistics in parentheses

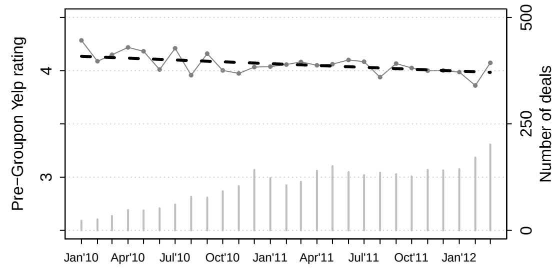
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



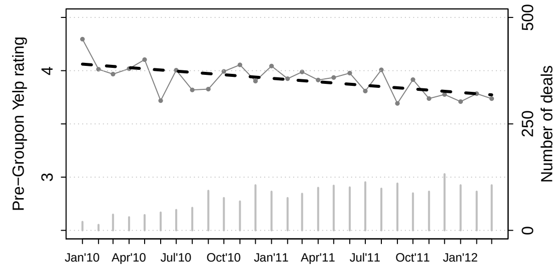
(a) New York



(b) Boston



(c) San Francisco



(d) Seattle

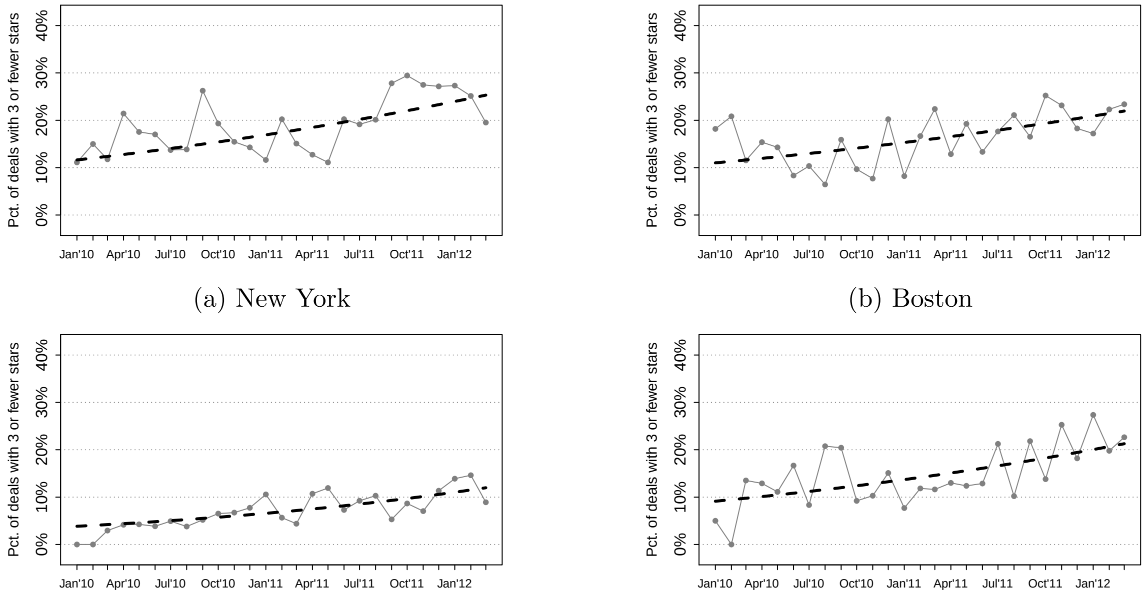
Figure 2: Month-over-month average Yelp rating (grey line) and number of Groupon deals (grey bars). The black dashed line shows a linear fit to the complete set of deals.

well as a logistic trend line. Low-rated deals appear to be increasing in frequency.

The trend line is computed by modeling the probability p_i that a deal is low-rated. Specifically, let Y_i be a Bernoulli random variable whose value is 1 a deal is low-rated, and 0 otherwise. Let $\mathbb{E}[Y_i|\tau_i] = p_i$. We then fit, by maximum likelihood, a logistic model with the following specification:

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1\tau_i. \tag{2}$$

As before, we first fit the model on the complete dataset, separately for each metropolitan area. The coefficient estimates are shown in the first part of Table 2. Their interpretation is in terms of additive changes to the log-odds of observing a low-rated deal. For example, looking at the coefficients for Boston, the odds of a low-rated deal on 1/1/2012 are $\exp(-2.09) = 0.124$, while 30 days later they are $\exp(-2.09+0.0312) = 0.128$. The corresponding predicted probabilities are approximately 11% and 11.3%. We then apply the same set robustness checks as before and report our results in Table 2. In all cases, we find a positive, statistically significant (at the 5%-level) association between the probability of observing a low-rated deal, and time. The magnitude of the effect varies little across different subsets of deals within each metropolitan area.



(a) New York

(b) Boston

(c) San Francisco

(d) Seattle

Figure 3: Month-over-month percentage of deals with an average Yelp rating of 3 or fewer stars. The black dashed line shows a logistic fit to the complete set of deals.

4 Discussion

Using simple regression analyses we have found a statistically significant, negative correlation between the time deals have been offered and the Yelp ratings of the merchants who offered them. A series of checks suggests the relationship is robust. Nevertheless, our finding is purely descriptive. Without understanding the underlying causes of this trend it is impossible to evaluate its implications for the daily deals business model. The change could be a result of a number of processes, each with distinct consequences. The following are some possibilities:

1. the population of merchants willing to run a Groupon deal remains, more or less, constant over time, but as Groupon is expanding the number of deals it offers, it has to work with some lower-rated merchants;

2. the population of merchants is changing, better rated are merchants dropping out of running Groupon deals, and Groupon has to substitute them for merchants with lower Yelp ratings who are offering the sames kinds of deals;

3. Yelp ratings are naturally eroding over time.

The consequences for Groupon depend on which, if any, of the above is the dominant process behind the observed trend.

Acknowledgements I thank Rocky Agrawal (`http://blog.agrawals.org`) for inspiring this work by suggesting that the quality of Groupon deals has been changing over time; John Byers, and Michael Mitzenmacher for their suggestions, and for fruitful earlier collaborations that laid the foundation for this work; and, Sharon Goldberg for her feedback on this draft.

References

Anderson, M., & Magruder, J. 2011. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal*.

Byers, John W., Mitzenmacher, Michael, & Zervas, Georgios. 2012a (February). Daily Deals: Prediction, Social Diffusion, and Reputational Ramifications. *In: The Fifth ACM WSDM Conference (WSDM'12)*.

Byers, John W., Mitzenmacher, Michael, & Zervas, Georgios. 2012b (June). The Groupon Effect on Yelp Ratings: A Root Cause Analysis. *In: Thirteenth ACM Conference on Electronic Commerce (EC'12)*.

Luca, M. 2011. Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School Working paper 12*, **16**, 12–016.